EXHIBIT 29

AEO

UNITED STATES DISTRICT COURT

FOR THE NORTHERN DISTRICT OF CALIFORNIA

SAN FRANCISCO DIVISION

RICHARD KADREY, et al.,

    Individual and Representative Plaintiffs,

  v.

META PLATFORMS, INC.,

          Defendant.

Case No. 3:23-cv-03417-VC

REBUTTAL REPORT OF

EMILY M. BENDER, PhD

FEBRUARY 03, 2025

AEO

83.   Dr. Ungar goes on to describe, at a high level, how the input texts are used in producing model weights through the process of training. At this point, the process becomes non-reversible, as he states (paragraph 132). However, this lack of reversibility does not negate the importance of original texts to the process of LLM development. The fact remains that the value that Meta found in the books used in the training data lies in the selection and arrangement of the words in the text. I elaborate on this point below.

84.   Dr. Ungar also argues that the processing done on the input data constitutes a "profound transformation" (paragraph 137). I note that Dr. Ungar does not provide a definition of "transformation" (or "transformativeness"), nor is he a legal scholar qualified to opine on the question of how the law sees the relationship between a set of training data and the weights derived from it.

85.   In another discussion of "transformativeness" (also paragraph 137), Dr. Ungar asserts that LLM outputs aren't based on the inputs. But if that were true, then constructing an LLM wouldn't require an enormous training data set. In short, there are three components of an LLM: its model architecture, its training regime, and its dataset. Meta announces that the Llama models are "open-source", but one of these three components is not released nor thoroughly described, and that's the training data. Furthermore, the models are valued for their outputs (the text they can be used to synthesize). In my opinion, this shows that the outputs are very much dependent on the inputs.

C.   **The Value of Books to Meta is Squarely in the Selection and Arrangement of the Works Contained Therein**

86.   There is also plenty of evidence in Dr. Ungar's report that the value of the books as training data is precisely the selection and arrangement of the words contained therein. Dr. Ungar writes that LLM's "capabilities instead derive from the structure of the network and the connections/weights, which through training, enable the network to recognize patterns and relationships in the data" (paragraph 30). Without the words as arranged by the authors of the texts, there would be no patterns or relationships for the model to represent.

27

AEO

87.   Dr. Ungar also describes a process of deduplication, which is important to the effective training of LLMs (paragraph 209). The importance of deduplication implies the value of many, varied, quality texts. If the choice and arrangement of words didn't matter, deduplication wouldn't matter, either, since repeated sequences would be as valuable as varied sequences.

88.   Dr. Ungar emphasizes the enormous amount of computing power being used to process the training data (paragraph 140). This is a lot of processing, to be sure, and it means that the resulting artifact is not just a database of the input data. However, it also shows that Meta is finding a lot of value in the input data. The processing would be meaningless without it.

89.   Finally, Dr. Ungar uses "noise texts" as a point of comparison in his experiments and motivates them as follows: "To provide additional comparison, a second set of models are further pretrained on "noise" texts—repetitive and nonsensical texts that do not help the model learn any meaningful information (e.g., repetitions of the word "the")" (paragraph 152 point E). Presupposed in this characterization is the idea that the high quality texts included in the training data are high quality because they contain not nonsensical strings, but words carefully and intentionally arranged by authors.

90.   As discussion in Section II, Part 1, the methodology of corpus linguistics underscores that the import of a text lies not just in its length, but in the word choice and grammatical structures it includes. These are the patterns that the LLMs are designed to represent and serve as the basis for their functionality in outputting text that mimics all of the genres in their training data. Without data that displays the patterns, no amount of training would result in an LLM that can represent the patterns.

IX.   CONCLUSION

91.   In conclusion, I find that the opinions that Dr. Ungar presents as unsupported for the following reasons:

AEO

92.   His report is threaded through with misleadingly anthropomorphizing language about LLMs. This presentation is misleading as to the functionality of LLMs, setting them up as something other than systems for representing the distribution of word forms in text.

93.   On the basis of this anthropomorphization, he develops a misleading presentation of the notion of "generalization", wherein the LLMs are agents doing something with their input data and this somehow means that the input texts were not material to the construction of the models.

94.   Dr. Ungar's choice of the MMLU benchmark rests on unsound scientific practice, as that benchmark does not measure what it purports to measure: language models do not understand anything, so a benchmark designed for language models to measure their "language understanding" is meaningless. Dr. Ungar furthermore doesn't establish any other foundation for the benchmark in order to make it meaningful in his experiments.

95.   Dr. Ungar's anthropomorphizing language displaces accountability. In using anthropomorphizing language, which places the Llama models as the agent of actions, Dr. Ungar displaces accountability away from Meta and to the models. In any deliberations about the development and use of "AI" systems, it is important to maintain clarity on who is taking action.

96.   Finally, Dr. Ungar's treatment of the value of written works to the development of large language models is internally inconsistent and unsupported. He argues both that "high quality" written works like books are necessary for LLM development, and (simultaneously) that any given written work is not important. These two positions are inconsistent. Even the largest dataset of written works is made up of individual works and cannot be created without them. Dr. Ungar also describes the processing of the input data as if it constituted "transformation". This does not change the fact that the value to Meta of the works was their original selection and arrangement of words.

_____

Emily M. Bender, Phd

29